Title: Auditory Forward Masking Explained by a Subcortical Model with Efferent Control of Cochlear Gain

Abbreviated Title: Efferent model for auditory forward masking

Authors: Braden N. Maxwell,[1] Afagh Farhadi,[3] Marc A. Brennan,[4] Adam Svec,[5] and Laurel H. Carney[1,2]

[1]Department of Biomedical Engineering, University of Rochester, Rochester, NY, 14642

[2]Department of Neuroscience, University of Rochester, Rochester, NY, 14642

[3]Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY, 14627

[4]Department of Special Education and Communication Disorders, University of Nebraska-Lincoln, Lincoln, NE, 68583

[5]Department of Audiology, San José State University, San José, CA, 95192

Corresponding author: Laurel Carney; Laurel_Carney@urmc.rochester.edu

Number of pages: 30

Number of figures: 6

Number of tables: 1

Number of words for abstract: 220

Number of words for introduction: 646

Number of words for discussion: 1256

Conflict of Interest Statement: The authors declare no competing financial interests.

Acknowledgements: Supported by NIH-NIDCD-R01-010813 (BNM, AF, LHC) and a grant from the Hearing Industry Research Consortium (RFP2022) (AS, MB). Braden Maxwell is currently affiliated with the Department of Psychology at the University of Minnesota. Afagh Farhadi is currently affiliated with the Department of Speech, Language, and Hearing Sciences at Purdue University.

Abstract:

Previous physiological and psychophysical studies have explored whether feedback to the cochlea from the efferent system influences forward masking. The present work proposes that the limited growth-of-masking (GOM) observed in auditory-nerve (AN) fibers may have been misunderstood; namely, that this limitation may be due to the influence of anesthesia on the efferent system. Building on the premise that the unanesthetized AN may exhibit GOM similar to more central nuclei, the present computational modeling study demonstrates that feedback from the medial olivocochlear (MOC) efferents may account for GOM observed physiologically in onset-type neurons in both the cochlear nucleus and inferior colliculus (IC), as well as in human psychophysics. Additionally, the computational model of MOC efferents used here generates a decrease in masking with longer masker-signal delays similar to that observed in IC physiology and in psychophysical studies. An advantage of this explanation over alternative physiological explanations (e.g., that forward masking requires inhibition from the superior paraolivary nucleus) is that this theory can explain forward masking observed in the brainstem, early in the ascending pathway. For explaining psychoacoustic results, one strength of this model is that it can account for the lack of elevation in thresholds observed when masker level is randomly varied from interval-to-interval, a result that is difficult to explain using the conventional temporal-window model of psychophysical forward masking.

Significance Statement:

The simulations presented here demonstrate that a recent computational model of the auditory subcortex including medial-olivocochlear efferents generates forward masking, an increase in detection threshold for a short probe tone following a preceding sound. This model explains results from both physiological recordings and psychoacoustic experiments. The theory that efferent control of cochlear gain is a contributing mechanism for forward masking has several advantages. This theory can explain masking exhibited by brainstem neurons, a phenomenon not explained by current physiological theories in which forward-masking originates later in the ascending pathway. Additionally, this theory explains results for a

psychoacoustic task with random variation in masker level, results not explained by the theory that persistent masker energy interferes with detection of the probe.

Introduction

Psychoacoustic forward masking is the elevation of detection threshold for a short probe tone due to a preceding sound. Psychophysical studies often assume that persistence of the neural masker response causes this increase in threshold (Jesteadt et al., 2005). However, physiological studies challenge this theory, showing reduction in neural probe responses (Nelson et al., 2009). The present study proposes medial-olivocochlear (MOC) efferent gain control as a mechanism for this reduction and suggests that MOC efferents can account for some aspects of psychophysical forward masking better than the classic temporal-window model of persistence (Moore et al., 1988).

Neural mechanisms hypothesized to explain decreased probe responses include adaptation of the synapse between the inner hair cell (IHC) and the auditory nerve (AN) (Smith, 1977; Oxenham, 2001) and inhibition of the inferior colliculus (IC) (Salimi et al., 2017; Gai, 2016). The efferent system has also been proposed (Jennings, 2021; Krull and Strickland, 2008; Relkin and Turner, 1988). However, physiological growth of masking (GOM, or increased probe threshold with increasing masker level) reported for the AN, which should show the effects of the efferent system, is insufficient to account for psychophysical GOM. Specifically, the range of forward-masked thresholds in anesthetized AN is only 15-20 dB (Relkin and Turner, 1988; Turner et al., 1994), whereas psychophysical thresholds for similar stimuli may increase by >40 dB across a masker-level range of >80 dB (Plack and Oxenham, 1998; Oxenham and Plack, 1997).

Observations in later neural stages, however, indicate GOM similar to that in psychophysics. Nelson et al. (2009) and Ingham et al. (2016) report some onset-type IC responses that exhibit GOM and neural thresholds with increasing time delay between masker and probe that are similar to psychophysical results. Nelson et al. (2009) noted that these patterns were generally not explained by the persistence of the masker response. Instead, the probe response itself was reduced.

Several studies have considered the possibility that probe suppression in the IC arises from the brainstem's superior paraolivary nucleus (SPON; Nelson et al., 2009; Felix et al., 2015; Gai, 2016; Gao and Berrebi, 2016; Gao et al., 2017; Salimi et al., 2017). However, chemical inactivation of the SPON does not remove suppression of probe responses for onset neurons, which demonstrate GOM similar to that in psychophysics (Felix et al., 2015, their Fig. 8A). Additionally, Ingham et al. (2016) found GOM in

cochlear nucleus (CN) onset neurons comparable to GOM in IC onset neurons, suggesting a potential neural source earlier than the midbrain.

The source of this suppression may be clarified by considering anesthetic effects. IC recordings (Nelson et al., 2009; Ingham et al., 2016) were made in awake and urethane-anesthetized animals; whereas AN recordings (Relkin and Turner, 1988; Turner et al., 1994) used a combination of barbiturate and urethane anesthesia (Dial in urethane). The results of Guitton et al. (2004) demonstrate that efferent activity is decreased more by barbiturate anesthesia than by urethane anesthesia, suggesting that differences in GOM observed in the AN versus IC may be attributable to different anesthetic states rather than different stages of auditory processing. Greater GOM might be observed in AN fibers if recordings were made using non-barbiturate anesthesia.

Consequently, reconsideration of the efferent system as an explanation for forward masking is warranted. We demonstrate that a subcortical model with MOC efferents (Farhadi et al., 2023) explains GOM in neurons that primarily respond to onsets in the IC and CN (Ingham et al., 2016). This model displays a realistic improvement in thresholds with increasing masker-signal delay, similar to both physiological and psychoacoustic studies. The model also accounts for psychophysical results from Jesteadt et al. (2005) that are difficult to explain using the conventional temporal-window model of persistent energy in forward masking (Moore et al., 1988, Oxenham & Plack 1998, Jesteadt et al., 2005). This investigation offers evidence for the efferent system as a mechanism underlying both physiological and psychoacoustic observations of forward masking.

Materials and Methods

Stimuli

The stimuli were tone forward maskers followed by short probe tone signals. Stimuli were matched to Ingham et al. (2016), Nelson et al. (2009), or Jesteadt et al. (2005) (Table 1). For Jesteadt et al. (2005), in the 0-σ rove condition, the masker was always 70 dB SPL. In the 6-σ rove condition, the masker level was randomly selected from a Gaussian distribution with a mean of 70 dB SPL and a standard deviation of 6 dB. The distribution was limited to +/- 3 standard deviations. While Ingham et al. (2016) and Nelson et al. (2009) matched the masker and probe frequencies to the best frequency of the

neuron, an IC neuron with a characteristic frequency (CF, the frequency to which a neuron is most sensitive) of 4 kHz was modeled here for simplicity. GOM modeling used the stimulus of Ingham et al. (2016) because that study reported GOM data across a range of masker levels for the onset cell-type specifically, whereas Nelson et al. (2009) reported data combined across cell types.

Table 1: Table of Stimulus Parameters. Parameters for the stimuli replicated from Ingham et al. (2016), Nelson et al. (2009), and Jesteadt et al. (2005).

Stimulus Parameters

Parameter	**Ingham et al. (2016)**	**Nelson et al. (2009)**	**Jesteadt et al. (2005)**
Masker Tone Duration	100 ms	200 ms	200 ms
Masker Level	Varied re Threshold	40 dB re Threshold	70 dB SPL; roved
Masker Ramps	2 ms	10 ms	2 ms
Probe Tone Duration	25 ms	20 ms	10 ms
Probe Ramps	2 ms	10 ms	5 ms
Masker-Probe Delay	2.83 ms	Varied 0-150 ms	0 and 30 ms
Masker Frequency	Varied, 4 kHz modeled	Varied, 4 kHz modeled	4 kHz in on-frequency condition; 2.4 kHz in off-frequency condition
Probe Frequency	Varied, 4 kHz modeled	Varied, 4 kHz modeled	4 kHz

Models

The modeling approach was similar to that used for predictions of forward-masked thresholds in Brennan et al. (2023). Two models of the AN were used to simulate neural responses with (Farhadi et al., 2023) and without (Zilany et al., 2014) gain control by MOC efferents. Both AN models had frequency selectivity based on measurements in humans (Ibrahim & Bruce, 2010; Shera et al., 2002) and a CF of 4 kHz. Time-varying AN responses provided the inputs to an IC model cell (Mao et al., 2013) that had a band-enhanced modulation transfer function (MTF, average discharge rate as a function of modulation frequency) with a best modulation frequency of 64 Hz, the median value reported for a population of IC band-enhanced neurons (Kim et al., 2020). Band-enhanced neurons, which were the model IC neurons in results presented here, have average discharge rates that are elevated over a band of modulation frequencies in comparison to the average discharge rate in response to an unmodulated stimulus. Masked tone-detection thresholds were calculated as described below based on the discharge rates of both model AN high-spontaneous-rate (HSR) fibers and the same-frequency inhibition-excitation (SFIE; Nelson and Carney, 2004) model for IC neurons. Note that the IC model of Mao et al. (2013) projected to the MOC, while the final model used for calculating IC thresholds was the SFIE model of Nelson et al. (2004). Mao et al. (2013) is a bandpass modulation filter that provides easily predictable responses to modulation and thus works well as a component of the efferent model. However, the Mao et al. (2013) simple filter model rings after stimulus offset, a property that is not typical in IC neurons. The SFIE model, which does not have this ringing property, was used for the final threshold calculations.

The MOC stage in the model with efferent control of cochlear gain (Farhadi et al., 2023) receives two inputs (Fig. 1): Input 1 is a wide-dynamic range (WDR) feedback signal that controls cochlear gain such that inner hair cells (IHCs) saturate in response to sound levels that are relatively high compared to the spectrum as a whole. This signal was simulated using a low-spontaneous-rate fiber (LSR) AN fiber; as the LSR fiber discharge rate increases, cochlear gain decreases. Input 2 is a fluctuation-driven signal from the IC that decreases cochlear gain for fluctuating channels.

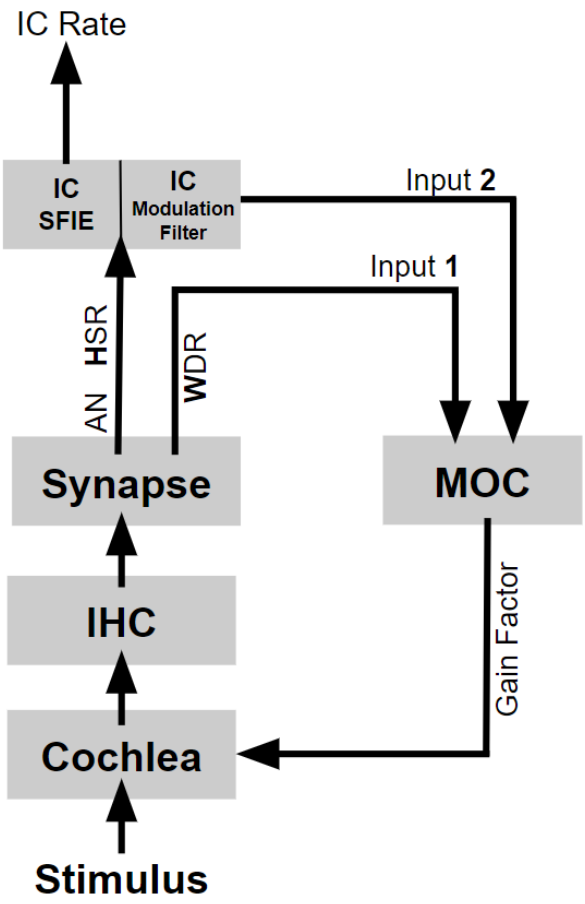

Figure 1: Schematic diagram of subcortical model with MOC Efferents (adapted from Brennan et al., 2023). Cochlear gain is decreased if the input level (and therefore the WDR discharge rate) is increased. Cochlear gain is also decreased if IC discharge rate is increased by fluctuations in HSR fiber responses. The IC modulation-filter model of Mao et al. (2013) projected to the MOC, while the final model used for calculating IC thresholds was the SFIE model (Nelson et al., 2004).

The process by which the efferent model produced GOM is shown in the blue curves of Fig. 2. Note that the HSR average discharge rate in response to the tone masker was similar during the 40- and 70-dB SPL maskers (left and right columns, respectively), because the average rate of the HSR fibers

saturated at or below 40 dB SPL. In contrast, the WDR average discharge rate during the masker was substantially greater at 70-dB SPL (right). In the MOC efferent model of Farhadi et al. (2023), higher WDR rates decrease cochlear gain. Therefore, the gain value decreased more during the 70-dB SPL masker (right) than during the 40-dB SPL masker (left). This difference in gain caused a greater decrease in average discharge rate during the probe when the probe followed the higher-level masker, a pattern that is visible in both the early (HSR and WDR) and more central (IC) responses (Fig. 2). For the model without efferents (gray), the probe response is still present (unmasked) following the 70-dB SPL masker. Note also that the feedback from the IC ("input 2" in Fig. 1) played a relatively small role in these simulations due to the lack of fluctuations in tone maskers.

Code Accessibility

Matlab code used for the simulations presented here will be made available at https://osf.io/ey4ht/.

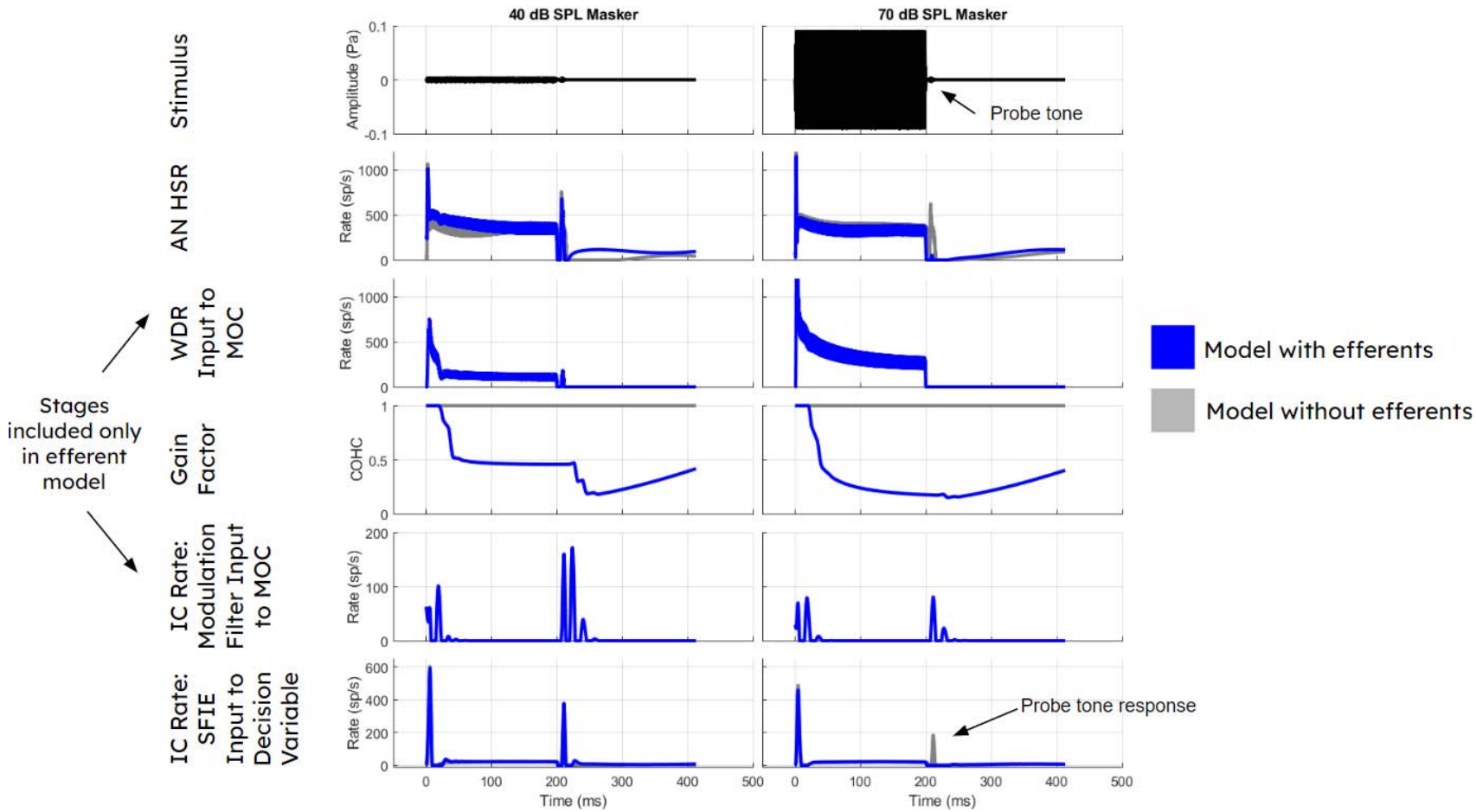

Figure 2: Impact of tone masker level on response to probe tone. The level of a 4-kHz tone masker (left and right columns, 40 and 70 dB SPL) influences the response at various stages of the efferent model (blue). Responses of model without efferents are shown for comparison (gray). Rows show stimulus and five stages of the model: HSR AN fibers, WDR responses (modeled using LSR AN fibers), the gain factor or influence of both feedback pathways on cochlear gain over time; the response of the fluctuation pathway (mostly irrelevant for tone maskers that

do not induce fluctuations); and the final response of the IC SFIE model. The WDR stage and feedback pathways are not included in the model without efferents. Note that, for the efferent model (blue), the HSR and IC responses to the probe tone (40 dB SPL, 10 ms, 2.8 ms delay) are substantially decreased for the higher-level masker in comparison to the probe tone responses for the lower-level masker condition, due to a change in cochlear gain. However, in the model without efferents (gray), the probe tone responses remain prominent for the 70-dB SPL masker.

Statistical Methods: Threshold Calculation

Estimated thresholds were based on simulations of a two-interval, forced-choice method-of-constant-stimuli paradigm. Model responses were used to decide, for each trial, whether the tone was in interval 1 or 2. For the physiological simulations, on each trial, the interval with the higher maximum response during a 40-ms time window was selected as the target interval. The use of maximum response during a time window to detect onset responses was based on the modeling approach of Brennan et al. (2023). The time window used for the physiological simulations matched the analysis windows used in Nelson et al. (2009) and Ingham et al. (2016), and is shown as a short solid horizontal line in Fig. 3. This window started 5 ms after the beginning of the probe tone stimulus so that the window included most of the probe response, while excluding the masker response, in line with the windows in Nelson et al. (2009) and Ingham et al. (2016). In most conditions, the masker and probe responses did not overlap in time, matching the lack of persistence of the masker response reported by Nelson et al. (2009).

For the psychophysical simulations, this procedure was the same except that the decision-variable window extended from 150 to 300 ms to reflect less a priori knowledge about the time of the probe response (Fig. 3: long dotted horizontal line). This window was not shifted in time for the different delays of 0 and 30 ms.

Percent correct was tabulated over 100 trials per target level. A logistic curve was fitted to the data across tone levels to estimate threshold (61% to match Ingham et al. (2016), 70.7% in all other cases). Note that the time windows used here differed from windows used in the temporal-window model of forward masking (Moore et al., 1988). Here, the windows were time spans over which response maxima were detected, rather than windows over which energy was integrated.

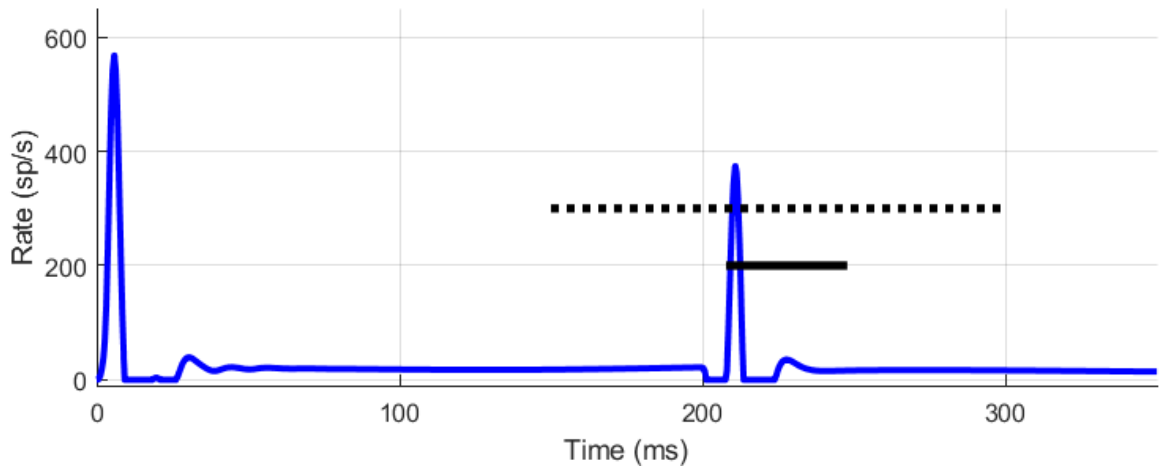

Figure 3: Decision-variable windows. IC SFIE model response to a 4 kHz, 40-dB SPL masker, 2.83 ms masker-probe delay, and 40 dB SPL, 10-ms probe tone (IC received input from model with efferents). Lines superimposed to show the placement of the physiological decision-variable window (bottom solid line) and psychophysical decision-variable window (top dotted line). The physiological decision-variable window was used for simulating the results of Ingham et al. (2016) and Nelson et al. (2009), and was shifted later or earlier depending on the masker-probe delay. The psychophysical decision-variable window was only used for simulating the paradigm of Jesteadt et al. (2005), and was not shifted with the masker-probe delay.

Specifying Thresholds in SPL vs. SL

Unit conversion was necessary for comparing the simulation results to physiological data because masker levels in the physiological papers (Ingham et al., 2016; Nelson et al., 2009) were described relative to neuron threshold (SL; suppression due to masker), whereas the model requires input to be scaled to a specific SPL. The following four points describe our approach: (1) For modeling the physiological data, we treated the unmasked threshold of the model neuron as analogous to the unmasked threshold of the neuron recorded in the experiment, and matched simulated masker levels relative to the unmasked threshold of the model neuron (SL). (2) For modeling the psychoacoustic task

with randomly varying level, masker levels in the simulation were matched to the SPLs used in the original experiment (Jesteadt et al., 2005). (3) It was possible to compare model thresholds to both psychoacoustic and physiological data because the model neuron unmasked probe tone thresholds were close to 0 dB SPL. Therefore, SPL and SL were roughly equivalent for the model. The greatest deviation from an unmasked threshold of 0 dB SPL was approximately 5 dB SPL for the AN model (with and without efferents) with the probe tone parameters used by Nelson et al. (2009). The IC model (with and without efferents) in the same condition had an unmasked threshold close to 1 dB SPL. (4) We looked for psychoacoustic data to compare to the model performance in the physiological tasks, to better understand whether one mechanism could explain both types of data. We chose data from a task that was as similar as possible to the physiological tasks (Moore and Glasberg, 1983). The masker-level conditions chosen for comparison to the model data were matched to the SPL used in the simulation. This choice allowed for realistic application of level-dependent model features, such as frequency selectivity and saturation of the synapse between the IHC and AN.

Although the points above describe the matching of masker level (simulation input) in the simulations, a separate question is whether to compare thresholds (simulation output) to physiological and psychoacoustic thresholds in terms of SPL or SL. It is reasonable to assume that some additional noise would elevate both unmasked and masked thresholds for the system as a whole, compared to a single-neuron model with the optimal CF for detecting the signal. Therefore, thresholds are expressed in SL (suppression due to masker) in all cases. In cases where psychoacoustic thresholds were expressed in SPL in the original paper without reporting unmasked thresholds, thresholds were converted from SPL to SL by subtracting 15 dB. This conversion was used by Oxenham and Plack (1998) for stimuli similar to those used here.

Results

Growth of Masking

Growth of masking in AN and IC Models with efferents (Fig. 4; blue squares, blue diamonds, respectively) was similar to physiological recordings of IC onset-type neurons from Ingham et al. (2016;

black solid line). The IC model with efferents inherited suppression properties from the AN model with efferents, as demonstrated by the similarity between IC and AN model results with efferents (Fig. 4: blue symbols). GOM was substantially greater for the models with efferents (blue) than the models without efferents (gray). The IC models generally had higher thresholds than the AN models. GOM was similar between (1) the models with efferents (blue symbols), (2) physiological recordings of onset-type neurons from Ingham et al. (2016; black solid line), and (3) psychophysical results reported by Moore and Glasberg (1983; dashed line) for similar probe tone duration, masker-probe delay, and SPLs similar to those used for the models here (note that the psychophysical data is matched to the model simulations based on SPL rather than SL; see Methods). Also note that CN onset-type neurons in Ingham et al. (2016) showed similar GOM to IC neurons, and thus a fit to the IC data implies a reasonable fit to the CN data, as the efferent GOM mechanism described here would have effects before the CN in the ascending pathway. However, the types of CN neurons observed in Ingham (2016) were not explicitly modeled here.

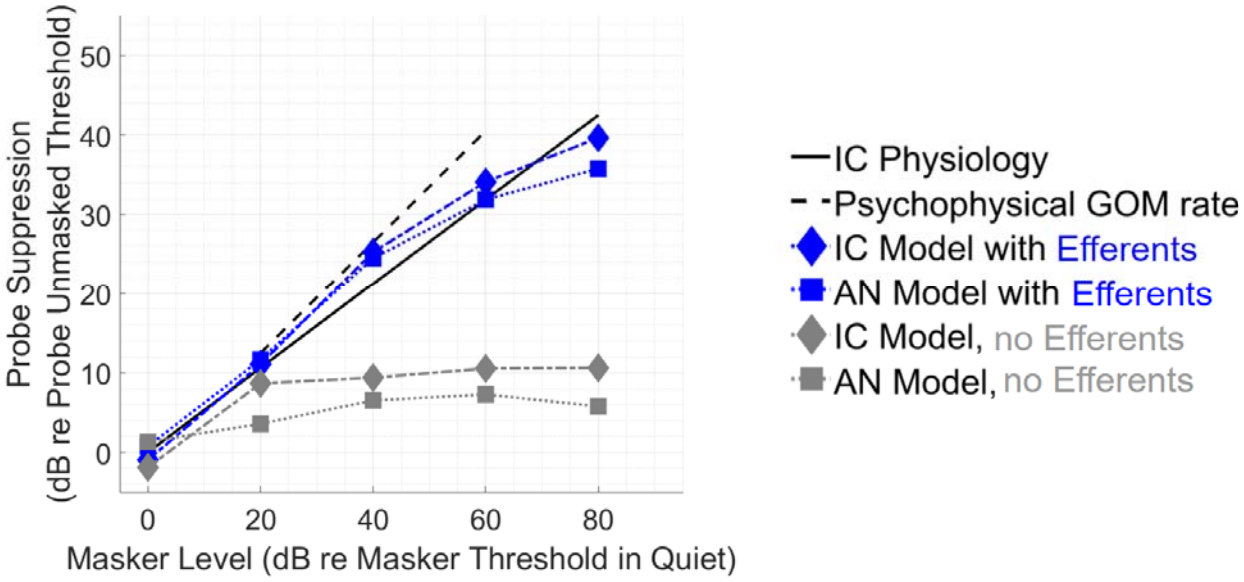

Figure 4: Forward-masked thresholds vs. masker level for models, physiology, and psychophysics. Growth of masking thresholds shown for IC and AN models (diamond and square markers, respectively), with and without efferents (blue and gray, respectively; color online). IC Physiology (black solid line) based on line fit to average suppression in onset-type neurons from Ingham

et al. (2016): 2-ms delay, 100-ms masker, 25-ms signal. Psychophysical GOM rate (slope of black dashed line) derived from Moore and Glasberg (1983): 0-ms delay, 210-ms masker, 20-ms signal. For physiology and models, masker level is plotted on the x axis relative to the threshold of the masker in quiet. Probe suppression is plotted on the y axis, in dB relative to the unmasked threshold of the probe. Psychophysical GOM curve is shifted along the x axis to more closely match the SPL of model stimuli (SL and SPL were similar for the model because model unmasked thresholds were within 5 dB of 0 dB SPL). The IC physiological results shown here (from Ingham et al, 2016) were from experiments using urethane anesthesia, with potentially less impact on the efferent system than barbiturates (Guitton et al., 2004).

Change in Masked Threshold with Masker-Probe Delay

In the physiological data of Nelson et al. (2009), masked threshold decreased by 19 dB for masker-signal delays between 0 and 150 ms (Fig. 5; black solid line). Without efferents, the AN model threshold decreased by 6 dB (gray squares), and the IC model threshold decreased by 11 dB (gray diamonds). Efferents brought this decrease in masked thresholds closer to the physiological range, leading to a decrease of 15 dB for the AN model (blue squares) and 19 dB for the IC model (blue diamonds). Suppression in the IC model with efferents was close to that of IC physiological recordings at short and long delays, but did not decrease as a function of delay rapidly enough at moderate delays. The IC model with efferents inherited most suppression properties from the AN model with efferents, although there were slight differences between the AN and IC. Relative to the AN, the IC sharpened the response to the probe at short delays because of sensitivity to modulation, and therefore the IC had lower thresholds than the AN in SPL (not shown; Fig. 5 shows suppression relative to unmasked threshold). This IC/AN difference applied for both models with and without efferents. However, the unmasked threshold of the AN was also about 4 dB higher than the IC (unmasked thresholds: IC with efferent: 1 dB SPL; IC without efferent: 0.8 dB SPL; AN with efferent: 5 dB SPL; AN without efferent: 5 dB SPL). This difference in unmasked thresholds caused suppression to be lower in the AN and obscured the enhanced encoding in SPL provided by the IC. Psychophysical and physiological data (dashed and solid black lines, respectively) were similar at short delays, and the IC model with efferents showed a similar decrease in threshold over these delays.

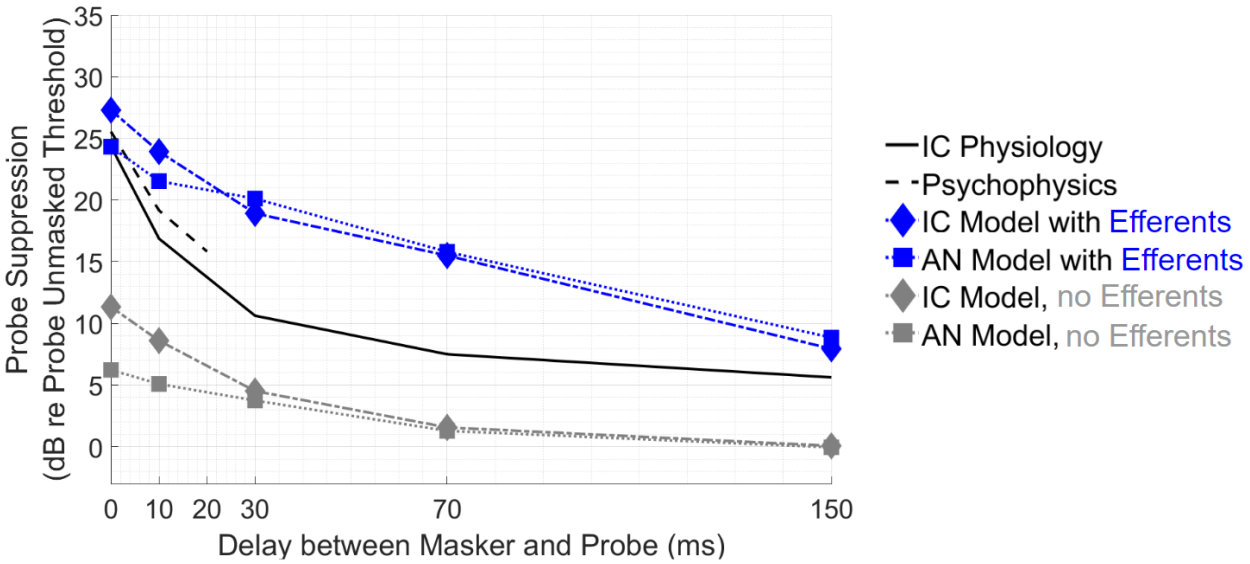

Figure 5: Change in Threshold with Increasing Masker/Probe Delay.

Suppression at varying delays shown for IC and AN models (diamond and square markers, respectively), with and without efferents (blue and gray, respectively). IC Physiology (black solid line): average suppression in awake recordings of IC neurons in Nelson et al. (2009): masker was 40 dB re neuron threshold (SL), 200-ms in duration, 20-ms signal. Psychophysics: from Moore and Glasberg (1983): masker at 40-dB SPL, 210-ms masker, 20-ms signal. Simulations matched parameters in Nelson et al. (2009); unmasked model thresholds were about 1 dB SPL for the IC models and about 5 dB SPL for the AN models, so 40-dB SPL and 40 dB SL were similar values for the models.

Masked threshold with roving-level masker

The previous results demonstrated the capacity of the Farhadi et al. (2024) IC model with efferents to account for aspects of both physiological and psychoacoustical data. However, to test this model head-to-head with the leading psychophysical explanation—that forward masking is primarily caused by persistent masker energy that interferes with detection of the probe—we also challenged the model with a psychophysical result from Jesteadt et al. (2005) that the temporal window model (Moore et al., 1988; Oxenham and Plack, 1998) cannot explain.

Jesteadt et al. (2005) reported that randomly varying (roving) the level of the masker between intervals did not increase thresholds, regardless of whether the masker was on-frequency (masker and probe at 4 kHz) or off-frequency (masker at 2.4 kHz, probe at 4 kHz). The on-frequency result is consistent with the temporal window model, as cochlear compression should cause the neural response to the masker to change little across the range of randomly varied levels when the masker is at the same frequency as the probe, and this consistent masker response should prevent the masker-level rove from impacting the detection task. However, the off-frequency result is contrary to the prediction of the temporal-window model, as the neural response to the masker should change with the roving level: according to that model, higher masker energy in the standard interval should cause listeners to mistake that interval for the target (a false alarm), and lower masker energy in the target interval should cause listeners to mistake that interval for the standard (a miss), thus increasing listener thresholds when the masker level is roved. Jesteadt et al. (2005) also observed that listener decisions in a roving-level forward-masking task were correlated with the masker level in the target interval but not the masker level in the standard interval, which again suggests that the false alarms and misses described above were not occurring in the way that the temporal-window model would predict. Although other models featuring persistence of masker energy, such as temporal template models (the temporal pattern of masker + probe response, stored in memory), can explain this result, the conventional temporal window model cannot (Jesteadt et al., 2005). Here we tested whether the present model could explain these results without a temporal template stored in memory.

For the following tests, the model decision process considered the response over a long duration, from halfway through the masker until the end of the stimulus interval. This window length differed from the short time windows used for modeling the physiological data, which were short to match the duration of time windows used to record spikes in the physiological studies (see Fig. 3). The window length was extended for the simulations of human psychoacoustics based on the assumption that human listeners may not be able to precisely attend to activity within the short time windows used to measure discharge rate in the physiological studies.

Thresholds for the on- and off-frequency maskers based on the IC model with efferents did not change when a roving masker level was introduced (pairs of blue open vs. blue filled symbols), matching

this difficult-to-explain feature of the psychophysical data (pairs of black open vs. black filled symbols). Even in the off-frequency masker condition, for which the temporal window model would predict a substantial rove effect, the IC model with efferents showed minimal effect of the rove, similar to the psychophysical data. Model thresholds were close to psychophysical thresholds in the on-frequency, 0-ms delay, and in the off-frequency, 30-ms delay conditions, but were elevated compared to psychophysical thresholds for the on-frequency, 30-ms delay condition, similar to Fig. 5. These thresholds reflect too much suppression in the model at that delay. The model also provides too little suppression at the 0-ms delay for the off-frequency masker (right panel).

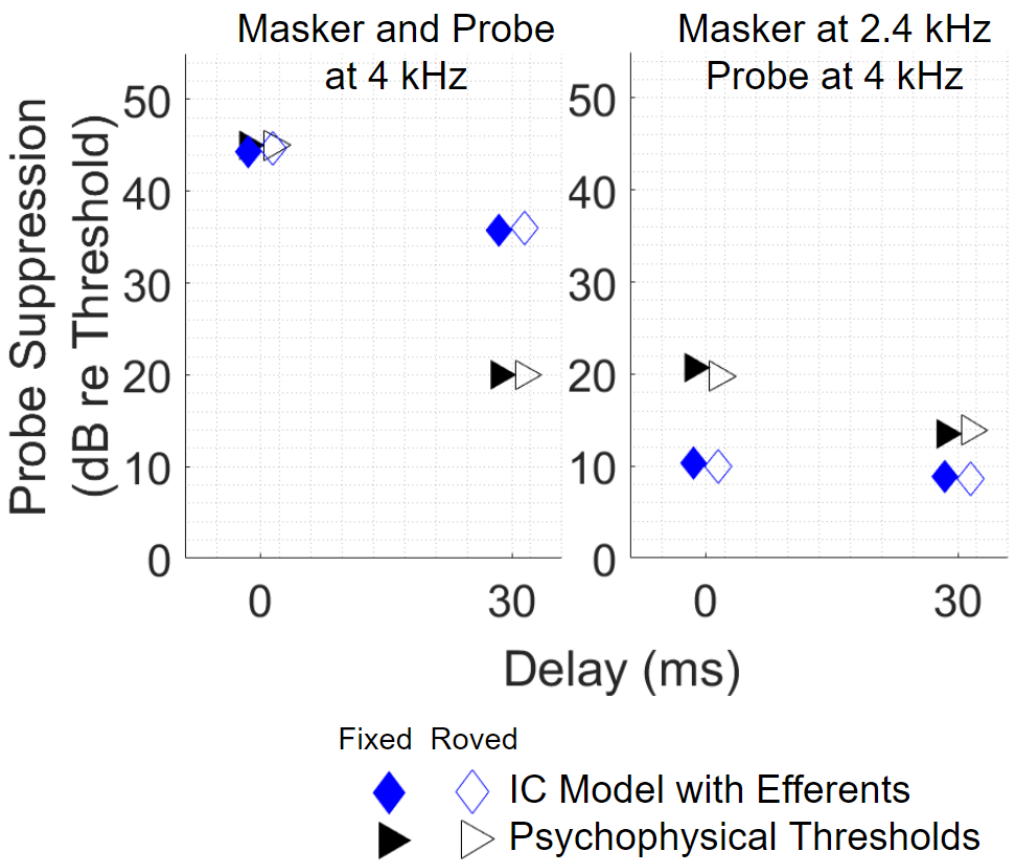

Figure 6: Stable Thresholds with Masker-level Rove. Psychophysical thresholds (black triangles; Jesteadt et al., 2005: experiment 1) and thresholds for IC model with efferent feedback (blue diamonds) shown for masker-probe delays of 0 and 30 ms, with masker level fixed at 70 dB SPL and with masker-

level rove (filled and open symbols, respectively). Left panel: on-frequency condition (masker and probe at 4 kHz). Right panel: off-frequency condition (masker at 2.4 kHz, probe at 4 kHz). Pairs of points with similar thresholds indicate the minimal impact of rove (x-coordinate offset to clearly show both filled and open symbols). Psychophysical suppression was calculated by subtracting a general estimate of unmasked thresholds for similar probe tones (15 dB SPL; Oxenham and Plack, 1998) from data reported in Jesteadt et al. (2005). Simulations matched parameters in Jesteadt et al. (2005).

This lack of rove effect (matching the psychophysical data) occurred because the IC model had minimal steady-state response to the tone masker (shown in Fig. 2), a property consistent with its use here to represent onset-type neurons. The false alarms and misses that would occur due to the roving masker level in the temporal-window model did not occur in this model because a higher masker level in the standard was not reflected in the steady-state response of the neuron, so a higher masker level would not be mistaken for a higher probe response. Likewise, the standard interval never caused a miss due to having a higher steady-state masker response than the target. Of course, the masker level in the target interval still influenced probe detection by producing changes in gain. Higher masker levels decreased gain, decreased the probe response, and decreased percent correct in some target intervals; lower masker levels increased gain, increased the probe response, and increased percent correct in some other target intervals. These cases balanced out (just as they would if the psychophysical task only involved one interval, instead of two) and yielded the same threshold as without rove, because the ratio of hits to false alarms was the same. Considering all of these factors together, model performance in this task depended only on the target interval, and not on the standard interval. This result aligns with the finding of Jesteadt et al. (2005) that the masker level in the target interval was correlated with whether trials were correct or incorrect, whereas the masker level in the standard interval was not. As mentioned above, Jesteadt et al. (2005) noted that their results were explainable if listeners matched a temporal template to the shape of the masker-probe response in the target interval. A temporal template and the IC decision process used here are similar in that they are robust to the roving-level paradigm. However, the present neural implementation has the advantage of not requiring any memory template.

It should be noted that thresholds for AN HSR model fibers (not shown) were *also* stable in the roving-level paradigm; however, this stability occurred for more complex reasons. In the case where masker and probe were both at 4 kHz, thresholds remained the same in the roving-level paradigm because the AN HSR fibers were saturated and changes in masker level were not reflected in the masker response. This result was consistent with the effects of compression as hypothesized by Jesteadt et al. (2005). In the off-frequency condition, in which the masker was at 2.4 kHz, the response to the masker steady-state did change with level; however, the change in the steady-state response to the masker between 70 and 88 dB SPL was small compared to the growth in the probe response near threshold (therefore, minimal false alarms occurred for trials with high masker levels in the standard interval), and the decrease in the steady-state response to the masker below 70 dB SPL did not create misses because the decision variable detected maximum discharge rate (and therefore always recorded the probe response, and not response to the masker, in those conditions). AN HSR responses to the off-frequency condition would not have been robust to the roving masker level if the simulated decision variable had been based on mean rates instead of maximum rate; however, the IC response would have still been robust to the roving level with either decision variable, because the IC model onset response contributed much more to the IC average discharge rate than did the IC steady-state response. Thus, although the masking itself was provided by the efferent feedback, the success of the IC model in simulating robustness to roving masker level was largely due to the IC model's lack of a steady-state response.

Discussion

Summary

The results above demonstrated that simulations based on a recent subcortical model (Farhadi et al., 2023) can explain several aspects of physiological and psychoacoustic tone forward-masking data. The MOC efferent feedback in this model added substantial GOM in the AN responses to a forward-masking paradigm, compared to a prior AN model (Zilany et al., 2014). This GOM was inherited by the IC model. The IC model had GOM similar to that in physiological recordings of onset-type cells in the IC (Nelson et al., 2009; Ingham et al., 2016), of onset-type cells in the CN (Ingham et al., 2016), and in human psychophysical results (Moore and Glasberg, 1983). This model had more forward masking

across masker-probe delay than the Zilany et al. (2014) model, which was based on AN responses under anesthesia. The model presented here also displayed a timecourse of changes in forward-masking thresholds that was similar to that in psychophysics (Moore and Glasberg, 1983) and IC physiology (Nelson et al., 2009). Finally, the lack of response to the steady-state of the masker in IC onset cells explained the stability of forward-masked thresholds in a paradigm with a randomly varying masker-level, a result that is difficult to explain using the temporal-window model of forward masking (Jesteadt et al., 2005).

Do the MOC efferents provide a plausible underlying mechanism for physiologically-observed forward masking?

Previous work suggested that physiological forward-masking varies depending on the response type of the neuron (e.g., onset, sustained, etc.) (Ingham et al., 2016). Similarities between physiological and psychophysical forward masking are strongest for onset-type neurons in the IC and CN. Physiological studies outline constraints for the underlying masking mechanism: (1) the mechanism may influence forward-masking at both CN and IC stages (Nelson et al., 2009; Ingham et al., 2016); (2) there is no evidence that the mechanism is reflected in the response of the barbiturate-anesthetized AN, perhaps due to suppression of the efferent system (Guitton et al., 2004, Relkin and Turner 1988; Turner et al., 1994), although the amount of forward masking present in the AN when the MOC efferent system is active is unknown; (3) the mechanism must decrease the response to the probe rather than temporally extending the response to the masker; that is, persistence of the masker response is rarely observed at the level of the IC (Nelson et al., 2009). Of course, the masker response may still persist, or overlap with the response to the probe, in later processing stages whose responses are shaped by slower time constants.

The mechanism of efferent-driven changes in gain fits all of these constraints: The efferent system is capable of influencing forward masking at both CN and IC stages because it influences cochlear gain; this system would be deactivated in the barbiturate-anesthetized AN, explaining the lack of physiological GOM observed by Relkin and Turner (1988; also Turner et al., 2004); and this system would decrease the response to the probe rather than temporally extending the response to the masker. Note

that forward masking observed in onset-cell types specifically requires probe suppression rather than persistence (regardless of whether psychophysical decisions are influenced by persistence), as these neurons have a minimal sustained response.

However, the proposed mechanism of MOC efferent-driven gain control raises important questions for future investigation: why, in the studies by Nelson et al. (2009) and Ingham et al. (2016), was there greater growth of masking for onset-cell types than other cell types, when efferent-driven cochlear gain control could hypothetically influence all cell types? Are the transient onset responses in the midbrain affected differently by efferents than steady-state responses? These questions merit further consideration.

Is it plausible that physiological forward masking, and specifically the MOC-efferent mechanism, can explain forward masking observed in psychoacoustic studies?

It is possible that the proposed efferent mechanism explains the forward masking observed in the IC and CN, but is not related to the masking observed in psychophysics. However, as an explanation for psychophysics, the efferent-gain mechanism has at least two significant strengths. First, as discussed above, the mechanism is suitable for explaining physiological results that bear striking similarities to psychophysical results. Figures 4 and 5 show several nontrivial similarities between physiological and psychophysical forward masking, including similar time courses and similar growth of masking.

Second, this model can explain the stability of thresholds with roving masker level in the results of Jesteadt et al. (2005), which pose a problem for the temporal window model of forward masking (Moore et al., 1988; Oxenham and Plack, 1998; Jesteadt et al., 2005). If the temporal-window model were correct, when the masker level was randomly varied, the persistent energy at the time of the probe should also vary, making the energy contributed by the probe more difficult to detect and increasing the thresholds. However, thresholds remain unchanged in this condition. The level of the masker must still be encoded, but variations in level do not interfere with detection of the probe.

The use of onset-type cells to encode the presence of the probe tone offers a solution to this problem. With minimal representation of the steady-state response of the masker, it did not matter whether the masker was roved up or down in level in the standard interval; the masker level did not cause

false alarms. Without interference from the standard interval, the cases of higher and lower gain in the target interval balanced out, leaving thresholds unchanged in the roving condition. This mechanism also explained why correct/incorrect answers by participants in Jesteadt et al. (2005) were correlated with masker energy in the target interval, but not with masker energy in the standard interval.

One question is why the cognitive algorithm for detecting the probe tone should use the output of onset neurons (which have greater GOM) to detect the probe, when there are other types of neurons with less GOM and therefore, potentially more salient probe responses at high masker levels. On the other hand, this disadvantage is outweighed by the fact that onset neurons do not encode the sustained response to the masker, and thus onset-type responses are pre-processed to focus on the relevant details for an algorithm that detects the beginning of a second sound (the probe). The Jesteadt et al. (2005) experiment is just one demonstration of how useful it is to ignore irrelevant aspects of the masker for the purpose of probe-tone detection.

Conclusions

Overall, the simulations presented here support the hypothesis that MOC efferents contribute to forward masking observed both physiologically (decreased probe responses observed in the IC) and psychophysically (increased detection thresholds for a probe tone following a masker). These simulations explained why listener thresholds are robust to randomly varying masker level, a surprising result that the temporal-window model (Moore et al., 1988) cannot predict. Additionally, this surprising result was explained without requiring templates stored in memory, instead using the simple mechanism of onset-type neurons that do not respond to the steady-state of tone maskers. These findings extend previous inquiries into the role of MOC efferents in psychophysical masking (Jennings and Strickland, 2012, 2010; Jennings et al., 2011; review: Jennings, 2021). The theory proposed here accounts for the GOM observed in recordings of onset-type neurons in the IC. Furthermore, this theory provides a conceptual account for how GOM could occur in the CN while not occurring in AN recordings performed under barbiturate anesthesia: the efferent system may be less active under that type of anesthesia compared to other anesthetics (Guitton et al., 2004). In summary, this theory provides potential answers to open

questions about forward masking stemming from previous experiments in both psychoacoustics and physiology.

References

Brennan MA, Svec A, Farhadi A, Maxwell BN, Carney LH (2023) Inherent envelope fluctuations in forward masking: Effects of age and hearing loss. The Journal of the Acoustical Society of America 153:1994–1994.

Farhadi A, Jennings SG, Strickland EA, Carney LH. (2023) Subcortical Auditory Model including Efferent Dynamic Gain Control with Inputs from Cochlear Nucleus and Inferior Colliculus. J. Acoust. Soc. Am. 154, 3644–3659. https://doi.org/10.1121/10.0022578

Felix RA, Magnusson AK, Berrebi AS (2015) The superior paraolivary nucleus shapes temporal response properties of neurons in the inferior colliculus. Brain Structure and Function 220:2639–2652.

Gai Y (2016) ON and OFF inhibition as mechanisms for forward masking in the inferior colliculus: a modeling study. Journal of Neurophysiology 115:2485–2500.

Gao F, Berrebi AS (2016) Forward masking in the medial nucleus of the trapezoid body of the rat. Brain Structure and Function 221:2303–2317.

Gao F, Kadner A, Felix RA, Chen L, Berrebi AS (2017) Forward masking in the superior paraolivary nucleus of the rat. Brain Structure and Function 222:365–379.

Guitton MJ, Avan P, Puel J-L, Bonfils P (2004) Medial olivocochlear efferent activity in awake guinea pigs. Neuroreport 15:1379–1382.

Ibrahim RA, Bruce IC (2010) Effects of peripheral tuning on the auditory nerve's representation of speech envelope and temporal fine structure cues. In, pp 429–438. Springer.

Ingham NJ, Itatani N, Bleeck S, Winter IM (2016) Enhancement of forward suppression begins in the ventral cochlear nucleus. brain research 1639:13–27.

Jennings SG (2021) The role of the medial olivocochlear reflex in psychophysical masking and intensity resolution in humans: A review. Journal of Neurophysiology 125:2279–2308.

Jennings SG, Heinz MG, Strickland EA (2011) Evaluating adaptation and olivocochlear efferent feedback as potential explanations of psychophysical overshoot. Journal of the Association for Research in Otolaryngology 12:345–360.

Jennings SG, Strickland EA (2010) The frequency selectivity of gain reduction masking: analysis using two equally-effective maskers. In, pp 47–58. Springer.

Jennings SG, Strickland EA (2012) Evaluating the effects of olivocochlear feedback on psychophysical measures of frequency selectivity. The Journal of the Acoustical Society of America 132:2483–2496.

Jennings SG, Strickland EA, Heinz MG (2009) Precursor effects on behavioral estimates of frequency selectivity and gain in forward masking. The Journal of the Acoustical Society of America 125:2172–2181.

Jesteadt W, Schairer KS, Neff DL (2005) Effect of variability in level on forward masking and on increment detection. The Journal of the Acoustical Society of America 118:325–337.

Kim DO, Carney L, Kuwada S (2020) Amplitude modulation transfer functions reveal opposing populations within both the inferior colliculus and medial geniculate body. Journal of Neurophysiology 124:1198–1215.

Krull V, Strickland EA (2008) The effect of a precursor on growth of forward masking. The Journal of the Acoustical Society of America 123:4352–4357.

Mao J, Vosoughi A, Carney LH (2013) Predictions of diotic tone-in-noise detection based on a nonlinear optimal combination of energy, envelope, and fine-structure cues. The Journal of the Acoustical Society of America 134:396–406.

Moore BC, Glasberg BR (1983) Growth of forward masking for sinusoidal and noise maskers as a function of signal delay; implications for suppression in noise. The Journal of the Acoustical Society of America 73:1249–1259.

Moore BC, Glasberg BR, Plack C, Biswas A (1988) The shape of the ear's temporal window. The Journal of the Acoustical Society of America 83:1102–1116.

Nelson PC, Carney LH (2004) A phenomenological model of peripheral and central neural responses to amplitude-modulated tones. The Journal of the Acoustical Society of America 116:2173–2186.

Nelson PC, Smith ZM, Young ED (2009) Wide-dynamic-range forward suppression in marmoset inferior colliculus neurons is generated centrally and accounts for perceptual masking. Journal of Neuroscience 29:2553–2562.

Oxenham AJ (2001) Forward masking: Adaptation or integration? The Journal of the Acoustical Society of America 109:732–741.

Oxenham AJ, Plack CJ (1997) A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing. The Journal of the Acoustical Society of America 101:3666–3675.

Plack CJ, Oxenham AJ (1998) Basilar-membrane nonlinearity and the growth of forward masking. The journal of the acoustical society of America 103:1598–1608.

Relkin EM, Turner CW (1988) A reexamination of forward masking in the auditory nerve. The Journal of the Acoustical Society of America 84:584–591.

Salimi N, Zilany MS, Carney LH (2017) Modeling responses in the superior paraolivary nucleus: implications for forward masking in the inferior colliculus. Journal of the Association for Research in Otolaryngology 18:441–456.

Shera CA, Guinan Jr JJ, Oxenham AJ (2002) Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. Proceedings of the National Academy of Sciences 99:3318–3323.

Smith RL (1977) Short-term adaptation in single auditory nerve fibers: some poststimulatory effects. Journal of Neurophysiology 40:1098–1111.

Turner CW, Relkin EM, Doucet J (1994) Psychophysical and physiological forward masking studies: Probe duration and rise□time effects. The Journal of the Acoustical Society of America 96:795–800.

Zilany MS, Bruce IC, Carney LH (2014) Updated parameters and expanded simulation options for a model of the auditory periphery. The Journal of the Acoustical Society of America 135:283–286.

Table Legends

Table 1: Table of Stimulus Parameters. Parameters for the stimuli replicated from Ingham et al. (2016), Nelson et al. (2009), and Jesteadt et al. (2005).

Figure Legends

Figure 1: Schematic diagram of subcortical model with MOC Efferents (adapted from Brennan et al., 2023). Cochlear gain is decreased if the input level (and therefore the WDR discharge rate) is increased. Cochlear gain is also decreased if IC discharge rate is increased by fluctuations in HSR fiber responses. The IC modulation-filter model of Mao et al. (2013) projected to the MOC, while the final model used for calculating IC thresholds was the SFIE model (Nelson et al., 2004).

Figure 2: Impact of tone masker level on response to probe tone. The level of a 4-kHz tone masker (left and right columns, 40 and 70 dB SPL) influences the response at various stages of the efferent model (blue). Responses of model without efferents are shown for comparison (gray). Rows show stimulus and five stages of the model: HSR AN fibers, WDR responses (modeled using LSR AN fibers), the gain factor or influence of both feedback pathways on cochlear gain over time; the response of the fluctuation pathway (mostly irrelevant for tone maskers that do not induce fluctuations); and the final

response of the IC SFIE model. The LSR stage and feedback pathways are not included in the model without efferents. Note that, for the efferent model (blue), the HSR and IC responses to the probe tone (40 dB SPL, 10 ms, 2.8 ms delay) are substantially decreased for the higher-level masker in comparison to the probe tone responses for the lower-level masker condition, due to a change in cochlear gain. However, for the 70-dB SPL masker, the probe tone responses remain prominent in the model without efferents [gray]).

Figure 3: Decision-variable windows. IC SFIE model response to a 4 kHz, 40-dB SPL masker, 2.83 ms masker-probe delay, and 40 dB SPL, 10-ms probe tone (IC received input from model with efferents). Lines superimposed to show the placement of the physiological decision-variable window (bottom solid line) and psychophysical decision-variable window (top dotted line). The physiological decision-variable window was used for simulating the results of Ingham et al. (2016) and Nelson et al. (2009), and was shifted later or earlier depending on the masker-probe delay. The psychophysical decision-variable window was only used for simulating the paradigm of Jesteadt et al. (2005), and was not shifted with the masker-probe delay.

Figure 4: Forward-masked thresholds vs. masker level for models, physiology, and psychophysics. Growth of masking thresholds shown for IC and AN models (diamond and square markers, respectively), with and without efferents (blue and gray, respectively; color online). IC Physiology (black solid line) based on line fit to average suppression in onset-type neurons from Ingham et al. (2016): 2-ms delay, 100-ms masker, 25-ms signal. Psychophysical GOM rate (slope of black dashed line) derived from Moore and Glasberg (1983): 0-ms delay, 210-ms masker, 20-ms signal. For physiology and models, masker level is plotted on the x axis relative to the threshold of the masker in quiet. Probe suppression is plotted on the y axis, in dB relative to the unmasked threshold of the probe. Psychophysical GOM curve is shifted along the x axis to more closely match the SPL of model stimuli (SL and SPL were similar for the model because model unmasked thresholds were within 5 dB of 0 dB SPL). The IC physiological results shown here (from Ingham et al, 2016) were from experiments using urethane anesthesia, with potentially less impact on the efferent system than barbiturates (Guitton et al., 2004).

Figure 5: Change in Threshold with Increasing Masker/Probe Delay.

Suppression at varying delays shown for IC and AN models (diamond and square markers, respectively), with and without efferents (blue and gray, respectively). IC Physiology (black solid line): average suppression in awake recordings of IC neurons in Nelson et al. (2009): masker was 40 dB re neuron threshold (SL), 200-ms in duration, 20-ms signal. Psychophysics: from Moore and Glasberg (1983): masker at 40-dB SPL, 210-ms masker, 20-ms signal. Simulations matched parameters in Nelson et al. (2009); unmasked model thresholds were about 1 dB SPL for the IC models and about 5 dB SPL for the AN models, so 40-dB SPL and 40 dB SL were similar values for the models.

Figure 6: Stable Thresholds with Masker-level Rove. Psychophysical thresholds (black triangles; Jesteadt et al., 2005: experiment 1) and thresholds for IC model with efferent feedback (blue diamonds) shown for masker-probe delays of 0 and 30 ms, with masker level fixed at 70 dB SPL and with masker-level rove (filled and open symbols, respectively). Left panel: on-frequency condition (masker and probe at 4 kHz). Right panel: off-frequency condition (masker at 2.4 kHz, probe at 4 kHz). Pairs of points with similar thresholds indicate the minimal impact of rove (x-coordinate offset to clearly show both filled and open symbols). Psychophysical suppression was calculated by subtracting a general estimate of unmasked thresholds for similar probe tones (15 dB SPL; Oxenham and Plack, 1998) from data reported in Jesteadt et al. (2005). Simulations matched parameters in Jesteadt et al. (2005).